



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2012

Optimal rules for patent races

Judd, Kenneth L ; Schmedders, Karl ; Yeltekin, Sevin

Abstract: There are two important rules to patent races: minimal accomplishment necessary to receive the patent and the allocation of the innovation benefits. We study the optimal combination of these rules. A planner, who cannot distinguish between competing firms in a multistage innovation race, chooses the patent rules by maximizing either consumer or social surplus. We show that efficiency cost of prizes is a key consideration. Races are undesirable only when efficiency costs are low, firms are similar, and social surplus is maximized. Otherwise, the optimal policy involves a race of nontrivial duration to spur innovation and filter out inferior innovators.

DOI: <https://doi.org/10.1111/j.1468-2354.2011.00670.x>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-60863>

Journal Article

Originally published at:

Judd, Kenneth L; Schmedders, Karl; Yeltekin, Sevin (2012). Optimal rules for patent races. *International Economic Review*, 53(1):23-52.

DOI: <https://doi.org/10.1111/j.1468-2354.2011.00670.x>

OPTIMAL RULES FOR PATENT RACES *

By Kenneth L. Judd, Karl Schmedders, and Şevin Yeltekin †

Hoover Institution, U.S.A.; University of Zurich, Switzerland.; Carnegie Mellon University, U.S.A.

November 18, 2010

Abstract

There are two important patent race rules: minimal accomplishment necessary to receive the patent and the allocation of the innovation benefits. We study the optimal combination of these rules. A planner, who cannot distinguish between competing firms in a multistage innovation race, chooses the patent rules by maximizing either consumer or social surplus. We show that efficiency cost of prizes is a key consideration. Races are undesirable only when efficiency costs are low, firms are similar, and social surplus is maximized. Otherwise, the optimal policy involves a race of nontrivial duration to spur innovation and filter out inferior innovators.

JEL: C61, C63, C73, L43, L50

Keywords: Patent Race, Patent Policy, Stochastic Dynamic Games, Markov-perfect Equilibria.

1 Introduction

Patent systems use races to spur innovation. Firms compete to be the first to develop a new product and obtain a patent and monopoly rights to sell that product. It's been long argued (see Wright (1983) for an earlier criticism) that patent systems may be suboptimal mechanisms because they spark races and generate wasteful duplication of effort. As Loury (1979) noted, races also have offsetting benefits: increased investment leads to quicker innovation. Therefore a patent system designed to encourage innovation must carefully weigh the benefits of quick innovation against the

*Manuscript received February 2010; revised October 2010.

†We are grateful to two anonymous referees and the editor Jan Eeckhout for helpful comments on an earlier version of this paper. We thank Ulrich Doraszelski, Mort Kamien, Ariel Pakes, Mark Satterthwaite, Christopher Sleet, and seminar audiences at Stanford University, MIT, Harvard University, the University of Texas at Austin, Northwestern University, Washington University in St. Louis, the University of Bielefeld, and at IESE/Barcelona for helpful discussions.

over-investment cost generated by races. In this paper, we provide a model of optimal patent rules that endogenizes the choice of races in designing incentives for innovation. Our environment has two important features. First, we assume that innovation is a multistage process and find that the waste induced by patent races can be moderated by filing requirements, that is, by deciding the stage of an innovation process at which the patent is granted. Second, we explicitly model the product market inefficiencies of a patent monopoly and argue that these inefficiencies must be considered along with any waste generated by a patent race when evaluating patent policies. In general, there are two ways to stimulate innovation: offer a big prize to a single innovator, or offer a smaller prize but use a race to threaten each firm that the prize will go to his competitor. When both incentive devices have inefficiencies, it is generally best to use both.

We model races as multistage stochastic games between heterogeneous firms. Firms differ in their cost of innovation. They proceed through several stages of progress (e.g. rough idea, blue print, prototype), with the final stage culminating in successful innovation and a marketable product. For a given patent policy, our dynamic, stochastic innovation race resembles those studied by Fudenberg et al. (1983) and Harris and Vickers (1985a,b, 1987). However, we endogenize patent policy by embedding this game into the problem of a patent authority that selects the rules governing the races.

We study the optimal policy of a patent authority who can verify partial success at the time of a patent application, but cannot observe an individual firm's efficiency. The patent authority has access to policy tools typically used in patent systems. First, it chooses when to award the patent, namely, it chooses the innovation stage at which a patent or an exclusive contract is awarded to a firm and the race is terminated. This represents the minimal accomplishment necessary and the filing requirements to obtain a patent and sets the length of a patent race. Firms race as long as no firm has achieved this level, but as soon as one has met the requirements of the patent rules, it proceeds with exclusive rights to develop the product. Second, the authority chooses how much of the benefits of an innovation go to the patent holder. This allocation is affected by rules such as patent length, patent breadth, and renewal fees.

Most analyses of patent policy focus on the optimal duration and breadth of patent protection¹, but assume that a firm does not receive a patent until its R&D process is complete². We distinguish between the stage a patent is granted and the end of the innovation process. This distinction allows us to evaluate the desirability of races and to analyze the effects of the race length on the pace and cost of innovation preceding and following a patent. This assumption is consistent with actual

¹For example, Gilbert and Shapiro (1990), O'Donoghue et al. (1998), Denicolo (1999), Hopenhayn and Mitchell (2001), Wright (1999), Green and Scotchmer (1995).

²Perry and Vincent (2002) is an exception. See Section 3 for a more thorough comparison of their setup and ours.

innovation processes. Firms often receive patents for a product in its early development stages, bear significant expenditures afterwards and reap financial benefits only when R&D is finished and the product is marketed. This feature is seen in many of the case histories of important inventions examined in Jewkes et al. (1969). For example, the first patents for xerography were granted many years before the first copy machine, and far more money was spent on development of the transistor after the patent was granted than before. When a patent should be granted represents a fundamental question in the design of patent policy. We show that this choice has important implications for the costs and benefits of innovation.

In our environment, the patent authority must weigh three considerations. These are the time at which R&D is completed and the product is ready, the potential waste of a patent race and the welfare loss in the product market from patent monopolies. If we make the conventional assumption that no patent is granted until all R&D is completed and choose a long life for the patent, then there will be much over-investment in the patent race and large welfare loss from the patent monopoly, but an early product introduction time. The waste in the patent race can be eliminated by an early grant of the patent or a smaller reward. Decreasing the value of the patent, by lowering patent length and breadth, will reduce the product market inefficiencies, and reduce the excessive investment activity, but may lead to poor intertemporal resource allocation in a multistage patent race. Additionally, it may be difficult to filter out the less efficient firm when patents are granted early; even firms with larger costs of innovation may find it feasible to compete for a few stages to obtain a valuable patent. A priori, it is not obvious which effect dominates in choosing the optimal policy.

We show that both policy instruments, when to award the patent and rewards to the winner, will generally be used to spur competition and innovation. In most circumstances, under reasonable assumptions about product markets, it is optimal for a social surplus maximizing patent authority to grant a patent after considerable progress has been made by the firms. In other words, races are desirable. They serve two important purposes in our model. First, the patent authority uses races to motivate innovators when the prize alone cannot, due to inefficiencies or limitations, provide adequate incentives. Second, a patent race serves as a filtering device. A race is used to increase the chance that the patent is rewarded to the most efficient innovator.

Another important factor that influences the optimal mix of the two policy instruments is the preferences of the patent authority. We consider two different specifications – social and consumer surplus – whereas most analyses in the patent literature focus only on social surplus. We examine optimal patent policy when the patent authority maximizes consumer surplus because it may represent the preferences of the median voter who is likely to be a consumer waiting for new goods. We show that consumer surplus maximizing patent authorities always prefer races, with or without

product market distortions.

Our choice of patent policy instruments is influenced by existent national patent institutions. National patent systems are applied uniformly to all inventions and use a small set of tools such as filing requirements, duration and scope of a patent, and renewal fees.³ Our focus is on the trade-offs between using a race versus wealth transfers to the winner, so we examine the interaction of two instruments: when to grant a patent and the value of the patent to the winner. The value of the patent incorporates other patent policy details such as breadth, duration and scope, therefore we do not model them separately. We also find that our insights are robust to the addition of other instruments. For example, we consider the possibility of using an auction to find the more efficient innovator and avoid excessive transfers to it. We show that when deadweight losses from patent monopolies and/or patent prizes are significant, races are still part of the optimal patent policy even when auctions are available. The key fact is that racing is a useful alternative to stimulating innovative effort when the use of large prizes as an incentive device is limited by their nontrivial inefficiency costs.

The paper is organized as follows. Section 2 discusses the relevant literature and elaborates on the differences between our approach and those contained in related papers. Section 3 lays out our model of a patent race, describes the patent authority's problem and presents our analytical results. Section 4 reports numerical results on the race dynamics. Section 5 displays the optimal patent policy and provides robustness checks. Section 6 considers the optimal patent policy under alternative informational assumptions. Section 7 concludes the paper with a discussion of possible extensions of the present analysis. All proofs and a detailed description of our numerical method are included in the Appendices.

2 Related Literature

This paper contributes to a growing literature that studies the implications of R&D competition on the design of optimal patent policy and helps bridge a gap between two distinct lines of research. The first of these lines focuses on R&D competition, taking patent policy as given. The second endogenizes patent policy, but largely abstracts from the R&D competition that precedes the award of the patent.

³Our model can also be used in a procurement context. The main distinction between patent systems and procurement problems is that in the latter case, the buyer can draw on a much larger set of instruments to give proper incentives to competitors. For example, the buyer could offer some payments to the loser and could require some interim reporting to monitor firms' progress. In procurement problems, each buyer can design an incentive system that is tailored for the particular product in question. Our focus in this paper is on national patent systems which use limited set of instruments.

Early contributions to the first line of research include Kamien and Schwartz (1982), Loury (1979), Lee and Wilde (1980), Reinganum (1981, 1982) and Dasgupta and Stiglitz (1980a,b). In these models, the probability that a firm successfully obtains a patent at each date depends only on the firm's current R&D expenditure and not on its past R&D experience. Competition takes place in "memoryless" or "Poisson" environments (see also the survey article by Reinganum (1989)). This first generation of models was subsequently extended by Fudenberg et al. (1983) and Harris and Vickers (1985a,b, 1987) to incorporate learning or experience effects. Throughout, patent policy was taken as given.

In contrast, contributions to the second line of research, including Nordhaus (1969), Klemperer (1990), Gilbert and Shapiro (1990), O'Donoghue, Scotchmer, Thisse (1998), Denicolo (1999, 2000), and Hopenhayn and Mitchell (2001), Llobet, Hopenhayn and Mitchell (2006), endogenize the policymaker's choice of patent length and/or breadth. However, these papers largely abstract from the R&D races that precede the award of a patent. Rather, they mainly focus on minimizing post-innovation distortions.

Our paper is at the intersection of these two lines of research. It is complementary to a growing body of work that studies the effects of patent policy on the strategic interaction between competing firms. This body of work includes Scotchmer and Green (1990, 1995), Perry and Vincent (2002) and Fershtman and Markovich (2010). The first paper by Scotchmer and Green focuses upon the novelty requirement and disclosure aspects of patent law and their implications for the pace of innovation. Their second paper involves the role of patent licensing in environments with successive innovations. Both papers consider a truncated R&D phase, with two stages of innovation only.

Fershtman and Markovich (2010) consider multistage races. Their main focus is on comparing "strong" and "weak" patent regimes. A strong patent regime entails patent protection at the end of the race and a no-imitation rule on intermediate discoveries. A weak patent regime allows for imitation of any discovery: intermediate or final. They study the effect of these two different regimes on firm innovation in a multistage race of a fixed length. We, on the other hand, distinguish between the stage at which a patent can be awarded and the stage at which an innovation is successfully completed. In Fershtman and Markovich (2010), races are exogenous and of fixed length. In our setting, they arise endogenously as a function of the optimal patent policy.

Perry and Vincent (2002) study the design of patent/procurement races when the planner cannot observe the firms' innovation stage. The focus is on how to induce the laggard participants to drop out without distorting other firms' incentives to invest. We examine similar issues, but our environment differs from theirs along a few important lines. First, the informational asymmetry in our model stems from the inability of the planner to observe the cost of investment (i.e. the type of firm), rather than their innovation state. Second, in our setting, the prize to the winner

of the patent, which incorporates patent length and breadth, is endogenously determined, whereas in theirs, it is a parameter. Third, we explicitly incorporate into our model externalities from monopoly distortions or deadweight loss of taxes used to finance them. This allows us to study the effectiveness of races in spurring innovation when the ability of the patent authority to use prizes is limited by these externalities.

Our patent instruments are different from those utilized in multiple innovation settings considered in O'Donoghue, Scotchmer and Thisse (1998) and Hopenhayn, Llobet, and Mitchell (2006). In O'Donoghue et al., the focus is on increasing profits for successive improvements in innovation while minimizing monopoly distortions. The patent authority is assumed to have several instruments that control the length and the breadth of the patent. The optimal mixture of these instruments is not identified, instead two types of policies, with different lengths and breadths are compared in terms of the effects they have on the diffusion and costs of innovation. In Hopenhayn et al., the focus is on optimal rewards for successive innovation and a system of buyouts and licensing that can implement these rewards. Again, the policy instruments are the length and the scope of the patent. Our model is about a single innovation, not a sequence of innovations. Our main emphasis is on the role of races. We analyze the innovation phase of a single good, and examine at what stage a race should be terminated and patent-type protection be given.

Patent races are not the only mechanism for spurring innovation. Research tournaments, where contestants compete to find the innovation with the highest value to the sponsor and receive a prespecified prize, can be and are used to achieve a similar goal. Research tournaments are particularly useful when research inputs are unobservable and research outcomes cannot be verified by courts. Taylor (1995), Fullerton and McAfee (1999) are amongst the papers that study such tournaments, the former in an environment with identical firms, the latter with firms of heterogeneous ability. Innovation races and research tournaments differ both in institutional and model details. In a single, well-defined innovation race, the quality requirement is fixed, the time of innovation is variable. The focus is on the pace of innovation and the competition between the firms. In a research tournament, quality is variable, the terminal date, on the other hand, is fixed. In research tournaments, the emphasis is on the quality of the product, not on the pace of innovation. In the McAfee-Fullerton model, an entry auction filters out less efficient firms, in our model, filtering is mostly achieved by varying the quality requirement, even when auctions are available. Our goal is to endogenize the choice of races and to study the changes in the innovation pace and intensity when quality requirements and prizes are chosen optimally. Thus we find patent races to be the most natural environments in which to achieve this goal.⁴

⁴See Scotchmer (1999) for an environment in which the patent system is optimal. She shows that if a direct mechanism cannot use ex-post information on value or costs, the only feasible incentive mechanisms are patent

3 A Model of Patent Policy for Multistage R&D Processes

We use a multistage stochastic innovation race model to evaluate patent policies. The introduction of a new product requires the completion of N stages of development by profit-maximizing firms that differ in their cost of R&D effort. We assume that each firm controls a separate innovation process. They have perfect information about each other's cost structure and position and choose their investment levels simultaneously. Each firm begins at stage 0 and the firm that first reaches the stage $D \leq N$ obtains exclusive rights to continue. The value of D corresponds to filing requirements for a patent. After winning the race, the patentholder completes the final $N - D$ stages without competition. When the patentholder reaches stage N , the patentholder markets the new product and earns the rents from a patent monopoly.

We use D to represent a firm's effort before it receives a patent relative to the total effort required to produce a marketable good. Our model calibrates costs so that D/N roughly represents the fraction of total expected cost incurred before a patent is granted. It is clear that D is neither 0 nor N in most patent systems: for most industries substantial expenditures occur before and after patents are granted. For example, pharmaceutical firms bear large R&D cost before receiving patents, but most also spend large sums on proving the safety and efficacy of any drug after receiving the patent. Our environment allows us to analyze the trade-off between extended monopoly power given to a patent recipient if $D = 0$ and the cost of duplicated effort in a long race if $D = N$. We also examine how the choice of D affects the other parameters of patent policy.

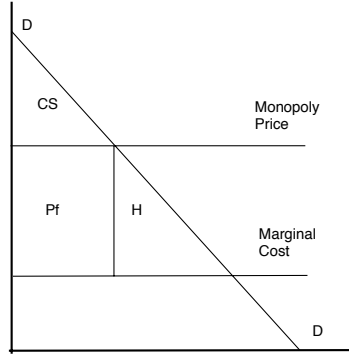
A patent granting authority (hereafter, PGA), who cannot continuously monitor all races and has imperfect information about the cost structures of firms involved, chooses the rules that govern races. Consistent with real patent systems, we assume that the PGA has only two policy tools: Ω , the prize to the patentholder, and D , the stage at which a patent is awarded. If $D = 0$, then there is no race. It also represents the case where the patent requirements are so minor that the patent goes to whomever, with trivial effort, first comes up with the barest notion of the innovation. In the game, it formally corresponds to the PGA giving the patent at random to one of the firms. The key assumption is that in this case each firm has equal chance of winning without having made any investment. The prize, on the other hand, may be literally a cash prize or, like a patent, it may be a grant of a monopoly which produces a profit flow with present value Ω . In the latter case, Ω is meant to represent many features of a patent. For example, Ω is small if the patent life is short or if it has small breadth, or if renewal fees are large. We tacitly assume that patent breadth, length, renewal fee rules that are associated with a specific Ω have already been determined by the PGA.

We next model the post-innovation market. We let B denote the potential value to society from

renewals systems with fees.

the invention. This includes the potential social surplus of a new good as well as any technological or knowledge spillover into other markets. The allocation of social benefits B is affected by patent policies. Figure 1 displays the per-period allocation before the patent has expired. Suppose that demand is given by DD and that there is a constant marginal cost of production. Figure 1 assumes that the patentholder can sell the new good at the monopoly price, but not engage in price discrimination, creating a profit Pf for the firm and leaving consumers with a surplus of CS . The area H represents the deadweight loss from monopoly pricing.

Figure 1: Division of Social Value



Once the patent has expired, the good is assumed to sell competitively at marginal cost, implying that consumers will receive all the social benefits, which equal $CS + Pf + H$. Profits from patents are proportional to demand, and, therefore, roughly proportional to social benefits B . Hence we assume that the prize to the patentholder is $\Omega = \gamma B$; it equals a proportion γ of the present value of potential social benefit.

In Figure 1, the deadweight loss H represents the social cost of monopoly profits in the patent system.⁵ More generally, we assume that the deadweight loss is proportional to the profits received by the innovator, and is equal to $\theta\Omega = \theta\gamma B$ for some $\theta \geq 0$. For example, $\theta = 0.5$ in Figure 1. This linear specification for the deadweight loss captures the basic point that $\gamma > 0$ causes inefficiencies, and is an exact description of this loss when demand is linear and marginal costs are constant, as well as, when demand has constant elasticity and marginal cost is zero. There

⁵Price controls may be used to reduce the deadweight loss, but they would also reduce monopoly profits and the prize. Long-lived patents increase γ but at the expense of increasing the total deadweight losses of monopoly. Cash prizes may be granted by the PGA along with shorter duration patents. This reduces the time during which the market experiences the deadweight loss H , but it only creates other inefficiencies since society bears the distortionary cost of the taxes used to finance the prize. Therefore, there are inefficiencies no matter what financing scheme is used.

are similar inefficiencies when Ω is a cash prize financed by distortionary taxes. In that case, θ represents the marginal efficiency cost of funds, a number which can plausibly be as low as 0.1 or as high as 1, depending on estimates of various elasticities, tax policy parameters, and the source of marginal funds; see Judd (1987) for a discussion of these factors. Therefore, the θ parameter represents either the relation between deadweight loss and profits for monopoly or the marginal efficiency cost of tax revenue.

While any patent is in effect, the firm receives profits, the consumers receive some benefit, but some of B is wasted in the transfer process. We assume that the patentholder's profits are γB but that the deadweight loss due to inefficiencies is $\theta\gamma B$, leaving consumers with $B - \gamma B - \theta\gamma B$.

We consider two different specifications of the PGA's preferences: social and consumer surplus. Most analyses assume that social surplus is the appropriate objective, but it is not clear that this should always hold. For example, in a procurement context one suspects that the PGA does not care about the profits of the participants in the race. Also, some argue that Congress should choose patent policy to maximize social surplus, but it is also imaginable that Congress will consider maximizing consumers' welfare if the median voter is more of a consumer than a producer. Social surplus, denoted by $W^S(D, \gamma; \theta, B)$, is equal to the present discounted value of the social benefit, B , minus the deadweight losses of transfers to the patentholder, $\theta\Omega$, and minus the total investment expenditures of all innovators in the patent race. Consumer surplus $W^C(D, \gamma; \theta, B)$, on the other hand, is equal to the present discounted value of the social benefit, B , minus the transfers to the patentholder and its associated deadweight loss: $(1 - \gamma)B - \theta\Omega$. The precise statements of the PGA's objectives are stated in Appendix B.3, along with our numerical method that solves for the optimal policy.

There may be ways other than a patent race to encourage R&D, but those alternatives typically cause inefficiencies. For example, a government could give a prize to an inventor, but then insist that it should be produced in a competitive market. This would avoid the monopoly inefficiencies of a patent, but at the cost of distortionary taxation to finance the prize. Similar concerns would apply to government financed research labs. Furthermore, most R&D takes place outside of government sponsored programs. Our analysis of the PGA's problem takes into account the inefficiencies of the PGA's choices.

3.1 The Firms: A Multistage Model of Racing

The patent race with a specific Ω and D creates a dynamic game between two firms.⁶ Let $x_{i,t}$ denote firm i 's stage at time t . We assume that each firm starts at stage 0; therefore, $x_{1,0} = x_{2,0} = 0$. If firm i is at stage n then it can either stay at n or advance to $n+1$, where the probability of jumping to $n+1$ depends on firm i 's investment, denoted $a_i \in A = [0, \bar{A}] \subset \mathbb{R}_+$.⁷ The upper bound \bar{A} on investment is chosen sufficiently large so that it never binds in equilibrium. Firm i 's state evolves according to

$$x_{i,t+1} = \begin{cases} x_{i,t}, & \text{with probability } p(x_{i,t}|a_{i,t}, x_{i,t}) \\ x_{i,t+1}, & \text{with probability } p(x_{i,t} + 1|a_{i,t}, x_{i,t}). \end{cases}$$

There are many functional forms we could use for $p(x|a, x)$. We choose a probability structure so that innovation resembles search and sampling. Let $F(x|x) = p(x|1, x)$, that is, $F(x|x)$ is the probability that there is no change in the state if $a = 1$. For general values of a we assume

$$(1) \quad \begin{aligned} p(x|a, x) &= F(x|x)^a \\ p(x+1|a, x) &= 1 - F(x|x)^a. \end{aligned}$$

This specification is analogous to hiring a people to work for one period and having them work independently on the problem of moving ahead one stage. While this specification is a special one, its simple statistical foundation helps us interpret our results.⁸

During R&D, firm i 's cost function is $C_i(a)$, $i = 1, 2$. It is assumed to be strictly increasing and weakly convex in a . For the remainder of the paper, we assume the cost function for firm i takes the following form

$$C_i(a) = c_i a^\eta, \quad \eta \geq 1, \quad c_i > 0, \quad i = 1, 2.$$

Firms discount future costs and revenues at the common rate of $\beta < 1$ and maximize their expected discounted payoffs.

⁶We focus on the duopoly case for reasons of tractability and ease of exposition. We also believe that the duopoly case can serve as a valid approximation for the monopolistically competitive markets where most innovations take place.

⁷We have computed solutions to our model with firms being able to advance more than one stage in each period. These changes do not lead to any results that contradict the basic insights of this paper. Computational results with larger jumps can be obtained from the authors upon request.

⁸This specification allows only forward movement. This is typical of most of the patent race literature, although exceptions are present. See Doraszelski (2003) for a model with "forgetting", that is, x_{t+1} may be less than x_t .

3.2 Equilibrium

The patent race involves two phases. When one of the firms reaches stage D , it is awarded the patent and becomes the only innovator. We refer to the subsequent innovation stages as the monopoly phase and denote it by $X^M = \{D, D+1, \dots, N\}$. Prior to the monopoly phase the position of the two firms is described by $x = (x_1, x_2)$. We refer to the set of states before the patent is granted as the duopoly phase and denote it by $X^D = \{(x_1, x_2) | x_i \in \{0, \dots, D\}, i = 1, 2\}$. Since we employ backward induction to solve for the equilibrium of the game, we first solve for the monopoly phase and then for the duopoly phase.

3.2.1 Monopoly Phase

Firm i precedes as a monopoly after it receives the patent. We formulate firm i 's monopoly problem recursively. At the terminal stage N , the innovation process is over and firm i receives a prize of Ω . In stages D through $N-1$, it spends resources on investment. Let $V_i^M(x_i)$ denote the value function of firm i if it is a monopoly in state x_i . V_i^M solves the Bellman equation

$$(2) \quad \begin{aligned} V_i^M(x_i) &= \max_{a_i \in A} \left\{ -C_i(a_i) + \beta \sum_{x'_i \geq x_i} p(x'_i | a_i, x_i) V_i^M(x'_i) \right\}, \quad D \leq x_i < N \\ V_i^M(N) &= \Omega. \end{aligned}$$

The policy function a_i^M of a firm i monopolist is defined by

$$(3) \quad a_i^M(x_i) = \arg \max_{a_i \in A} \left\{ -C_i(a_i) + \beta \sum_{x'_i \geq x_i} p(x'_i | a_i, x_i) V_i^M(x'_i) \right\}, \quad D \leq x_i < N.$$

Proposition 1. *Firm i 's monopoly problem at state $x_i \in \{0, 1, \dots, N\}$ has a unique optimal solution $a_i^M(x_i)$. The value function V_i^M and the policy function a_i^M are nondecreasing in the state x_i .*

Proof. See Appendix A. □

3.2.2 Duopoly Phase

We formulate the competition between the firms before stage D as a duopoly game. In the analysis of this game we restrict attention to Markov strategies. A pure Markov strategy $\sigma_i : X^D \rightarrow A$ for firm i is a mapping from the state space X to its investment set A . We define the firms' value

functions recursively. Let $\mathbb{V}_i(x)$ represent the value of firm i 's value function if the two firms are in state $x = (x_1, x_2) \in X^D$. We use the conventional notation that x_{-i} (a_{-i}) denote the state (action) of firm i 's opponent. If at least one of the firms has reached the patent stage D , firm i 's value function is defined as follows:

$$(4) \quad \mathbb{V}_i(x) = \begin{cases} V_i^M(x_i), & \text{for } x_{-i} < x_i = D \\ V_i^M(x_i)/2, & \text{for } x_i = x_{-i} = D \\ 0, & \text{for } x_i < x_{-i} = D. \end{cases}$$

If neither firm has received the patent, the Bellman equations for the two firms are defined by

$$(5) \quad \mathbb{V}_i(x) = \max_{a_i \in A} \left\{ -C_i(a_i) + \beta \sum_{x'_i, x'_{-i}} p(x'_i|a_i, x_i) p(x'_{-i}|a_{-i}, x_{-i}) \mathbb{V}_i(x'_i, x'_{-i}) \right\}, \quad x_1, x_2 < D.$$

The optimal strategy functions of the firms must satisfy

$$(6) \quad \sigma_i(x) = \arg \max_{a_i \in A} \left\{ -C_i(a_i) + \beta \sum_{x'_i, x'_{-i}} p(x'_i|a_i, x_i) p(x'_{-i}|a_{-i}, x_{-i}) \mathbb{V}_i(x'_i, x'_{-i}) \right\}, \quad x_1, x_2 < D.$$

We now define the Markov perfect equilibrium of the race.

Definition 1. A Markov perfect equilibrium (MPE) is a pair of value functions \mathbb{V}_i , $i = 1, 2$, and a pair of strategy functions σ_i^* , $i = 1, 2$, such that

1. Given σ_{-i}^* , the value function \mathbb{V}_i solves the Bellman equation (5), $i = 1, 2$.
2. Given the value functions \mathbb{V}_i , and the strategy function of his opponent, the strategy function σ_i^* for player i solves equation (6), $i = 1, 2$.

A Markov perfect equilibrium always exists.

Theorem 1. There exists a Markov perfect equilibrium.

Proof. See Appendix A. □

3.3 A Special Case: Linear Technology, Identical Firms, No Deadweight Loss

Before we proceed with the analysis of our general model, we discuss a simple, canonical example with linear technology, identical firms, and no deadweight loss associated with the prize. In this

special case, which serves as a useful benchmark, there is obviously no value to a race since the PGA's problem can be perfectly internalized in a firm's profit maximizing strategy. If the PGA awards the project to one firm with the full social value as the prize before any of the firms start investing (i.e. $D = 0$), then the firm's profits equal social surplus and the firm chooses the socially optimal investment level. A race would speed up innovation but lead to excessive investment.

Proposition 2. *If costs are identical and linear ($c = c_2/c_1 = 1$ and $\eta = 1$), transfers to the patent winner have no deadweight loss ($\theta = 0$), and the PGA maximizes social surplus, then the optimal policy sets $\gamma = 1$ and $D = 0$, implying no race.*

Proposition 2 serves as a useful benchmark, but relies on the assumption that monopolies have no efficiency costs or that taxes paid to the innovator have no distortionary cost, in addition to assuming homogeneous firms with a linear technology. If firms are heterogeneous, the planner may prefer a race to ensure that the more efficient firm is not eliminated by bad luck. If firms have (strictly) convex costs, the planner may also prefer a race since two firms can achieve a rate of total innovation more cheaply than each can do individually. As we have stressed in the introduction, we are interested in examining optimal patent rules when inefficiencies and distortions are present, and when firms are heterogeneous. These more important and relevant cases do not admit closed form solutions, unless drastic simplifications such as linear cost and linear probability transition functions are used. However, the resulting linear model would have equilibria with only boundary solutions such as zero effort or deterministic transition. We do not find such equilibria to be good descriptions of the uncertainty that firms face in research and development or representative of reasonable firm interaction in the context of a race. Therefore, we do not dwell on these extreme environments, but instead use numerical methods to analyze our more general model.

4 Race and Innovation Dynamics

The purpose of this section is to obtain a good understanding of firms' behavior in our model. We begin our analysis by examining the monopolist inventor. Next we analyze the interaction of firms in a race and how firms' effort levels change with the conditions of the race, governed by the PGA's policy choices.

4.1 Monopolist Inventor

Proposition 1 states that both the value function V_i^M and the policy function a_i^M of the monopolist are nondecreasing in the state x_i , that is,

$$V_i^M(x_i) \leq V_i^M(x_i + 1), \quad a_i^M(x_i) \leq a_i^M(x_i + 1) .$$

Using the same approach as in the proof of Proposition 1, we can also show that both the value and the policy functions are nonincreasing in the cost coefficient c_i of the cost function $C_i(a) = c_i a^\eta$, $\eta \geq 1$, that is,

$$\frac{\partial V_i^M(x_i)}{\partial c_i} \leq 0, \quad \frac{\partial a_i^M(x_i)}{\partial c_i} \leq 0 .$$

We illustrate these analytical results for a monopolist with a discount factor $\beta = 0.996$, a cost elasticity $\eta = 1.5$, and a transition probability of unit investment $F(x|x) = F = 0.5$. The firm must pass $N = 5$ stages of innovation to obtain a prize derived from the social benefit $B = 100$. A model with these parameter values serves as a benchmark throughout this section.

Figures 2 and 3 show the effort levels $a_i^M(x_i)$ and the values $V_i^M(x_i)$ for $x_i = 0, 1, 2, 3, 4$, as a function of $c_i \in [1, 10]$, respectively. The prize proportion, γ , is set to 1 resulting in a patent prize $\Omega = \gamma B = 100$. The highest curve is for $x_i = 4$, the second highest for $x_i = 3$ and so on. The bottom curve shows the respective values for the initial state $x_i = 0$. As predicted by the analytical results, the closer the monopolist inventor comes to the completion of the product, the higher are its value and effort. Similarly, both the value and the effort level decrease in the cost coefficient c_i .

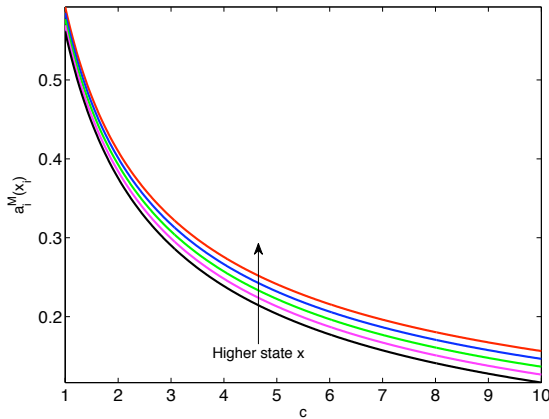


Figure 2: Monopoly Effort

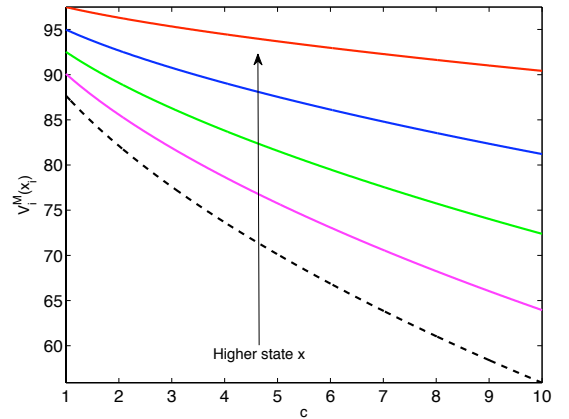


Figure 3: Monopoly Value

These simple observations have an important implication for the design of optimal patent policy. The PGA has the ability to award the project to one firm before any of the firms start investing. That is, the PGA can set $D = 0$ and randomly award the patent to one firm which then acts as a monopolist inventor. If the firms have identical cost functions, then it does not matter which firm receives the patent. But if one firm possesses a more efficient technology and has lower cost, then the PGA would prefer to give the patent to this firm. However, the PGA cannot distinguish between the firms, hence a choice of $D = 0$ does not allow for the selection of the more efficient firm. In fact, there is a 50 percent chance that the less efficient firm will receive the patent. Therefore, setting $D = 0$ becomes increasingly unattractive to the PGA as the cost heterogeneity among the competitors increases. To illustrate this effect, consider two firms with cost coefficients $c_1 = 1$ and $c_2 = 5$. If the more efficient Firm 1 receives the patent and gets to act as a monopolist inventor, then it obtains a value of $V_1^M(0) = 87.7$ and shows an effort of $a_1^M(0) = 0.561$ in state 0. If the less efficient Firm 2 receives the patent, then its corresponding values are considerably smaller, $V_2^M(0) = 70.1$ and $a_2^M(0) = 0.204$. The higher cost and lower effort of Firm 2 lead simultaneously to higher total cost of innovation (e.g., $c_1 a_1^M(0) < c_2 a_2^M(0)$) and to slower innovation (e.g., $1 - F^{a_1^M(0)} > 1 - F^{a_2^M(0)}$, see expression (1)). Both effects then result in the much smaller value $V_2^M(0)$. The expected value of a random patent award (which would also equal the expected producer surplus and social surplus) is then $(V_1^M(0) + V_2^M(0))/2 = 78.9$. Clearly the random award leads to a considerably smaller social surplus than an award to the efficient firm.

The numerical illustration of the analytical results regarding the monopolist explains why $D = 0$ may be an attractive policy choice for the PGA when firms have (nearly) identical cost, but becomes a less attractive option as the cost heterogeneity increases.

4.2 Duopolist Inventors

We now study the dynamic competition between the firms in a patent race and how the PGA's patent policy choices affect this competition. We first show the investment choices of the firms for a fixed patent policy. We then discuss the effect of the different policy choices on the firms' behavior along the innovation path. The firms' response to these choices provide intuition for the tradeoffs the PGA faces in setting the patent policy. The complexity of this response highlights the need for a computational approach for analyzing our model.

Throughout the following discussion the social benefit is set to $B = 100$ and the cost coefficient of Firm 1 is $c_1 = 1$ as in the benchmark case.

4.2.1 Firms' Investment Choices

The first set of graphs depicts the firms' investment choices along the innovation path of states 0 to 4 for two different values of $c = c_2/c_1 = c_2$: $c = 1$ and $c = 4$. The patent policy is fixed at $D = N$ and $\gamma = 1$, that is, the patent is only given after the completion of the final product and the prize to the patentholder is the entire social benefit, B . Figure 4 displays the investment effort of Firm 2 when firms' cost functions are identical. There are several notable features. First, the individual level of effort is more than 5 times that of the monopolist inventor's shown in Figure 2. Second, effort is higher when there is close competition, i.e. when firms are in identical or close states. Third, along the diagonal, firms increase their effort as they approach the patent award stage D . Thus the investment levels are highest in state $(4, 4)$. Finally, even in a symmetric cost race, when one firm gains a sufficient lead, the other firm effectively exits the race by reducing its effort close to 0 as apparent in states $(3, 1)$ and $(4, 1)$.⁹

Figure 5 shows the investment effort of Firm 2 when $c = 4$. Like the symmetric case, Firm 2 effort is at its lowest when Firm 1 takes a substantial lead. It is at its highest when Firm 2 is one step ahead of Firm 1, close to D . Intuitively, the less efficient firm has a higher chance of winning the patent award if it is close to the "finish line" D and it's ahead. In such a situation, Firm 2 increases its effort.¹⁰

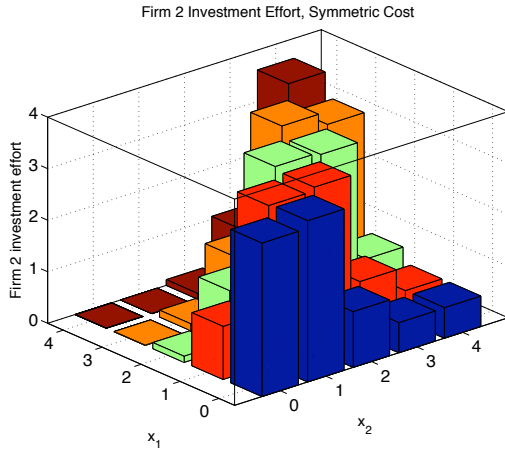


Figure 4: Symmetric cost: $c = 1$

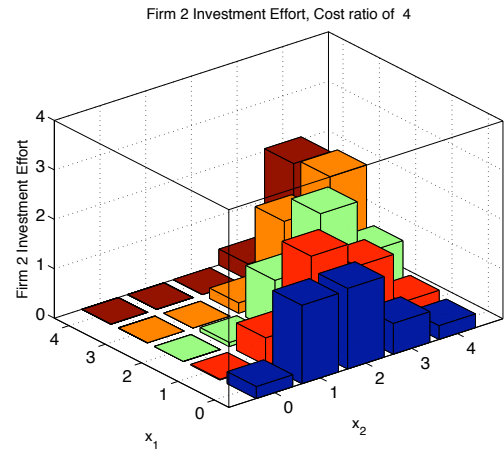


Figure 5: Asymmetric cost: $c = 4$

The previous graphs show the dynamics of firm investment for the fixed patent policy, $\gamma =$

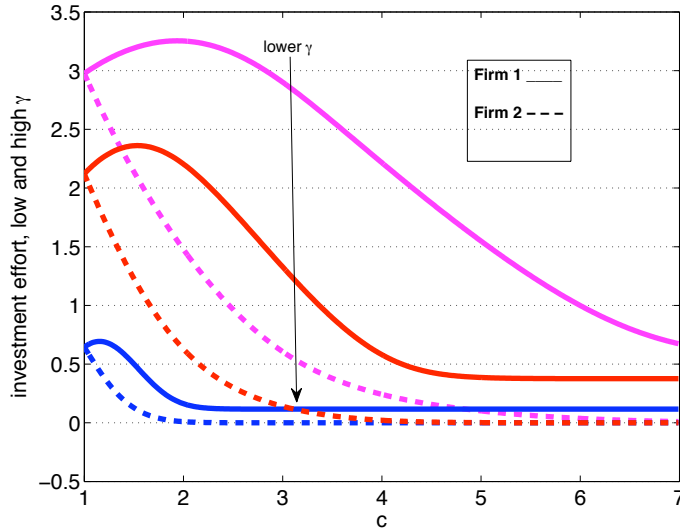
⁹ This is the famous "preemption" result prevalent in the patent race literature.

¹⁰ For c larger than 4, Firm 2 would need a larger lead close to D to have a good chance of winning the race. In those situations, Firm 2 effort is at its highest level when the firm is 2 or 3 steps ahead of Firm 1 and close to D . These results are omitted for space considerations, but can be obtained from the authors.

1, $D = N = 5$. We now investigate the effect of a change in the policy on firm effort and social surplus. The first result concerns the effect of the patent prize, γ , on the firms' investment effort, displayed in Figure 6. The patent award stage D continues to be fixed at N . The investment effort shown is for state $(0,0)$, the beginning of the race. The dashed lines represent Firm 2 effort and the solid lines represent Firm 1 effort. Figure 6 shows that regardless of the value of c , investment effort is decreasing in γ . Firms respond to a lower patent prize by investing less to reduce their cost. The response of firms' investment effort to γ is the same for all other symmetric states of innovation and for all race lengths, D .¹¹

Figure 6 also shows that firms lower their investment effort in state $(0,0)$ as c becomes sufficiently high. At these levels of c , Firm 2 finds it difficult to compete with Firm 1 and effectively drops out of the race. Thus, the race dissolves into a monopolist innovation process by Firm 1.

Figure 6: Duopoly Investment levels, for $D = N = 5$ and $\gamma \in \{0.1, 0.5, 1.0\}$



4.2.2 PGA's Policies and Innovation Dynamics

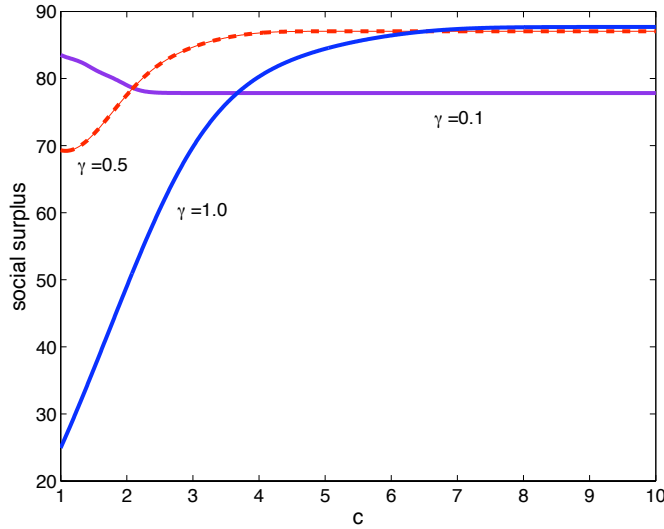
Based on the previous results, we have argued that firm investment is monotonic in γ for all D , for all cost asymmetry c , and state of innovation (x_1, x_2) . The effect of γ on social surplus is more complex. Figure 7 shows the social surplus for a variety of γ and c with D fixed at N . It is clear that social surplus is not always monotonic in γ . If firms have similar costs, they compete fiercely

¹¹Figures showing effect of γ on investment have been omitted from the text to conserve space, but are available from the authors.

(as shown in Figure 4), hence a large prize $\Omega = \gamma B$ leads to excessive investment and lower social surplus. As γ decreases, firms reduce their investment efforts. Although the resulting cost savings are slightly offset by slower innovation, the net impact is an increase in social surplus.¹² As a result, for small c values and $D > 0$, the PGA would prefer to reduce investment effort by setting a low prize, γ .

As c increases further, it becomes more expensive for Firm 2 to invest. Reducing γ for these values of c exacerbates this problem. By exerting lower effort, Firm 2 lessens the competition Firm 1 faces. Firm 1 responds by decreasing its own effort, innovation slows down, and social surplus decreases as a result. (See Figure 7 for $c = 3$). At these levels of c , an intermediate level of γ is preferred by the PGA because it restores some competition and faster innovation, without leading to massive duplication of effort. Firm 1 invests more, but Firm 2 effectively exits the race as soon as it falls behind.

Figure 7: Social Surplus for $D=N=5$ and $\gamma \in \{0.1, 0.5, 1.0\}$

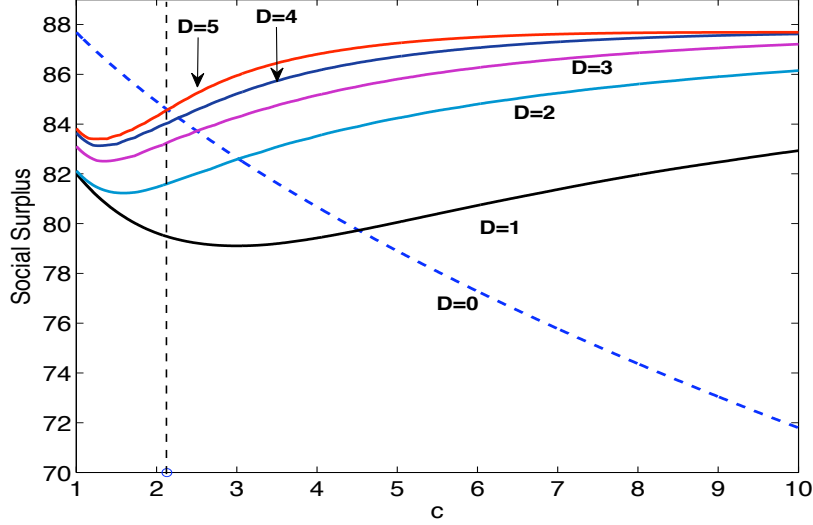


As c continues to rise (See Figure 7 for $c > 6$), Firm 2 drops out of the race for any prize γ , allowing Firm 1 to continue as a monopolist. To ensure that Firm 1 completes the innovation in a timely manner, a higher γ , in particular $\gamma = 1$ as in the monopolist inventor case, is needed.

The impact of the interaction between c and γ on social surplus, displayed by Figure 7 for $D = 5$, is qualitatively the same for all $D > 0$. The question remains, given the two instruments,

¹² Obviously, as γ tends to zero the firms stop investing and social surplus also tends to zero. Thus, social surplus is not monotone in γ .

Figure 8: Social Surplus



γ and D , what is the policy that maximizes social surplus? Figure 8 shows the social surplus for optimally chosen γ , for different D . The maximized social surplus is the upper envelope of the six lines. The optimal policy has a bang-bang feature. It involves no races until c reaches 2.13 (marked by the vertical dashed line) and full-length races thereafter. When firms are similar, they naturally compete fiercely, hence allowing them to race for a patent leads to excessive investment. Therefore until $c = 2.13$, the PGA prefers to flip a coin for the patent and let the winner invest without competition. At these low values of c , expected social surplus from a random patent assignment is higher than expected social surplus from a race with over-investment.

As c increases, Firm 2 finds it too costly to compete strongly, which leads to reduced effort, which in turn, results in less over-investment during a race. At the same time, the loss in social surplus associated with picking the inefficient firm at $D = 0$, increases. Therefore once c becomes large enough ($c \geq 2.14$ in Figure 8), a race is preferred to no race. More specifically, a race with $D = N$ delivers a higher social surplus than a race of any other length. The reason for this result is the following. The PGA prefers timely innovation by the efficient firm. Spurring innovation can be achieved two ways: one, by increasing the prize, γ and two, by decreasing D so that firms compete fiercely in the early stages of the race. However, lowering D increases the chances of the inefficient firm to win the race. Therefore the PGA uses a long race to filter out the inefficient firm. It uses the prize γ to fine-tune firm effort, so that the product is available in a timely manner without excessive over-investment. Table 1 shows how the social-surplus maximizing γ changes with c .

Table 1: Optimal policy choice γ conditional on c , for $D = 5$

c	1	1.5	2	3	4	6	8	10
γ	0.12	0.16	0.23	0.35	0.48	0.72	0.92	1.00

Table 2: Parameter Values

$\beta = 0.996$	discount factor
$N = 5$	number of stages of innovation
$B = 100$	total social benefit
$\eta \in \{1, 1.5\}$	elasticity of cost
$\theta \in \{0, 0.1, 0.25, 0.4, 1.0\}$	deadweight loss parameter
$c \in \{1, 1.25, \dots, 4.75, 5, 6, \dots, 20\}$	ratio of firms' costs coefficients, c_2/c_1 . ($c_1 = 1$)
$F(x x) = 0.5$	transition probability for unit investment

Our results so far have focused on the firms' reaction to different PGA choices of γ and D in the absence of product market distortions. We now turn our attention to optimal patent policies when the inefficiency costs of prizes are positive. We also consider optimal patent policies under consumer surplus maximization and under alternative innovation technologies.

5 Optimal Patent Rules

We now show optimal patent rule results when rewards have inefficiencies, firms are heterogeneous and technologies are allowed to be non-linear. Table 2 displays the set of parameters for our reported results.

These parameter values represent a wide range of cases. We make two normalizations: $c_1 = 1$ and $F(x|x) = .5$. We report detailed results from a 5-stage race only. Races with more stages do not provide any additional insights or change our qualitative results. The θ values are motivated by inefficiency costs of monopoly for standard demand curves and by the excess burden results in Judd (1987). The value of B is chosen so that races are neither too short nor too long. In general, the parameter values in Table 2 are chosen to represent innovation processes lasting from several months to a few years.

5.1 Social Surplus Maximizing Patent Rules

We now examine the case of social surplus maximization and the impact of cost heterogeneity and deadweight losses on optimal patent rules. Throughout this and subsequent sections, we use the term short (long) race to describe a patent awarded at a lower (higher) innovation stage. The terms slow and quick are used to describe the *time* it takes for firms to complete the innovation.

We first examine the effect of deadweight losses on optimal patent rules. To isolate this effect, we assume linear and homogeneous costs, as in Proposition 2. In this case, the only reason for having a race and bearing the inefficiency costs of duplicated effort is to reduce the deadweight loss associated with the patent prize. Figure 9 displays the social surplus as the deadweight loss parameter θ increases. The $D = 0$ line displays social welfare when one of the two identically efficient firms receives the patent in stage 0 and the prize γ is set to its optimal value. The $D = 4$ line displays social welfare when the patent goes to the first firm to reach that stage. Figure 9 shows that when θ is small, the planner prefers innovation with a monopolist as in the special case studied in Proposition 2. When θ exceeds 0.25, however, the planner switches to using a race. Note that $\theta = 0.25$ is a small value for deadweight losses; for a monopolist with a linear demand curve, $\theta = 0.50$. To see why the planner switches to a race as θ increases, consider the impact of θ on the optimal prize γ^* . It is clear from Figure 9 that for a fixed D , γ^* decreases as θ increases: larger distortionary costs lead to smaller prizes. Smaller prizes reduce monopolist investment levels and hence the speed of innovation. Therefore as θ increases, the PGA prefers races and the competition they spur to stimulate innovation.

We next add firm heterogeneity, and consider both linear and strictly convex costs. Table 3 reports, for a variety of θ values, the optimal prize and D , and the associated social surplus for different cost asymmetries, in both linear and convex cost cases.

The case with $\eta = 1.5$ and $\theta = 0$ is the case displayed in Figure 8 and discussed in detail in Section 4.2.2. For the other cases, there are multiple important results to note.

1. Social surplus, W^S , is decreasing in θ .
2. The optimal prize γ^* is weakly decreasing in θ for a fixed cost asymmetry c .
3. In the linear cost case, for c high enough, all positive D deliver the same social surplus.
4. For $\theta \geq 0.25$, regardless of the c value, the optimal policy always involves races, both in linear and in convex cost environments.

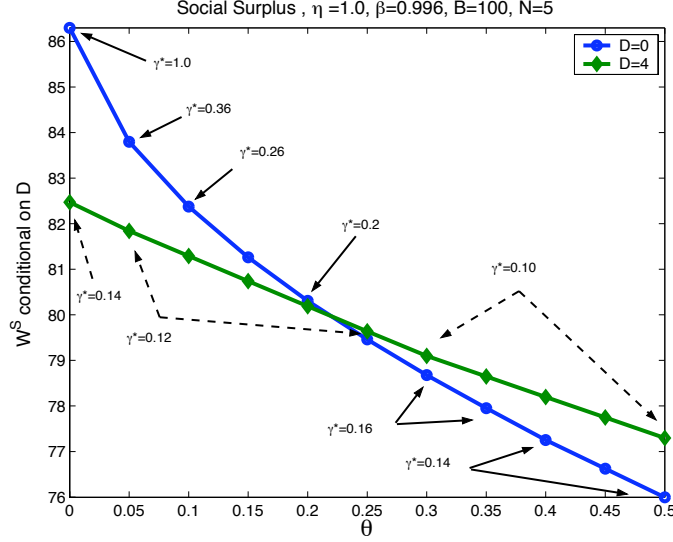
The intuition behind results (1) and (2) is straightforward, given the PGA's preferences. Result (3) is linked to the only difference between the linear and non-linear cost cases. In the linear cost

Table 3: Optimal Patent Policy for $B = 100, \beta = 0.996$

		$\eta = 1.5$			$\eta = 1.0$		
θ	c	D^*	γ^*	W^S	D^*	γ^*	W^S
0	1.0	0	1.00	87.7	0	1.00	86.3
	1.2	0	1.00	87.1	0	1.00	85.4
	1.5	0	1.00	86.2	0	1.00	84.0
	1.75	0	1.00	85.5	5	0.22	84.5
	2.0	0	1.00	84.9	5	0.28	85.0
	2.5	5	0.28	85.2	5	0.36	85.5
	4.0	5	0.48	86.8	5	0.66	86.2
0.1	1.0	0	0.34	83.0	0	0.26	82.4
	1.2	5	0.12	82.3	5	0.14	81.7
	1.5	5	0.16	82.1	5	0.18	82.3
	1.75	5	0.18	82.2	5	0.22	82.5
	2.0	5	0.20	82.4	5	0.26	82.5
	2.5	5	0.24	82.9	3	0.26	82.4
	4.0	4	0.32	83.0	*	0.26	82.4
0.25	1.0	5	0.10	81.3	3	0.12	79.7
	1.2	5	0.12	80.7	5	0.12	79.8
	1.5	5	0.14	80.2	3	0.16	79.9
	1.75	4	0.14	80.0	4	0.20	79.6
	2.0	4	0.16	79.9	2	0.20	79.6
	2.5	4	0.18	79.9	*	0.18	79.5
	4.0	3	0.22	79.7	*	0.18	79.5
0.4	1.0	5	0.10	79.9	4	0.10	78.2
	1.2	5	0.10	79.3	5	0.12	78.2
	1.5	4	0.12	78.6	2	0.14	77.9
	1.75	4	0.14	78.2	*	0.14	77.3
	2.0	3	0.12	78.0	*	0.14	77.3
	2.5	3	0.14	77.8	*	0.14	77.3
	4.0	2	0.16	77.2	*	0.14	77.3
1.0	1.0	5	0.08	75.5	4	0.10	72.8
	1.2	5	0.08	74.6	1	0.10	71.9
	1.5	3	0.08	73.3	2	0.12	70.5
	1.75	3	0.10	72.7	*	0.12	70.5
	2.0	3	0.10	72.3	*	0.12	70.5
	2.5	2	0.10	71.6	*	0.12	70.5
	4.0	2	0.12	70.6	*	0.12	70.5

* indicates that W^S is the same for all $D > 0$

Figure 9: Policy tradeoffs with linear and identical costs.



case, when c is high enough, Firm 2 exits the race. As long as there is a race, all D values deliver the same social surplus because Firm 1 proceeds as a monopolist inventor from the beginning. In the convex cost case, Firm 2's effort is always positive, no matter how disadvantaged it is compared to Firm 1. Therefore, even though the race turns into an “effective” monopoly invention process at high c , social surplus slightly varies with D and the indifference result does not hold.

The intuition behind result (4) is as follows. The presence of deadweight loss associated with prizes constrains the ability of the PGA to use them effectively to spur innovation. For example, observe that for $\theta \geq 0.25$, the optimal γ never exceeds 0.2 in Table 3. The constraint on γ forces the PGA to use D as an instrument to motivate competition instead. At high c values, when a low γ is not enough to induce Firm 2 to remain as a viable threat to Firm 1, a lower D pushes both firms to increase investment effort, by raising Firm 2's chances to win the race. However, if D is too low, then Firm 2 may win the race. The optimal choice of D represents a careful balance between filtering on the one hand, and fast innovation on the other hand.¹³

Our results so far indicate that if deadweight losses associated with monopolies or prizes are not included in the analysis, optimal patent rules display a bang-bang feature: they involve a sudden switch from no race to a full-length race as firm diversity increases (See the switch from $D^* = 0$ to $D^* = 5$ in the $\theta = 0$ case, both for $\eta = 1.0$ and $\eta = 1.5$). In contrast, when deadweight losses

¹³ The reason the convex cost case features longer races relative to the linear cost case has to do with effort levels less than one. At these effort levels, innovation costs are smaller and hence races are longer in the convex cost case.

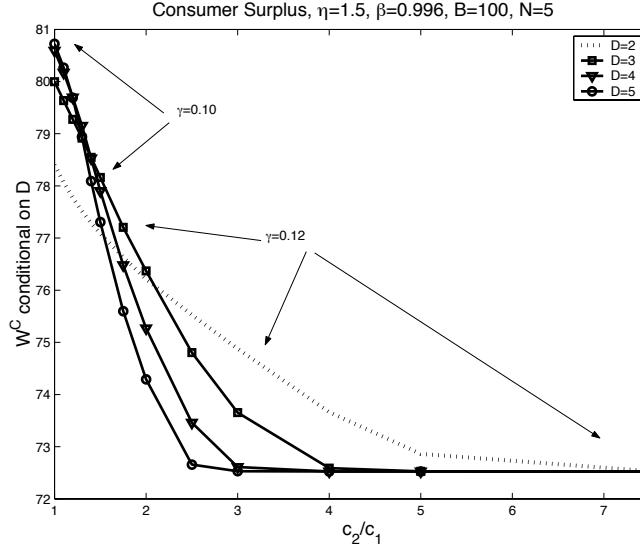
are explicitly considered, and they are of a reasonable magnitude, the optimal patent rule never involves a coin flip, it always features a race. The PGA prefers to grant patents at middle stages of development and give small prizes (See patent policy for $c > 1.5$ and $\theta \geq 0.25$). This suggests that patents should be easier to obtain, but less valuable.

5.2 Consumer Surplus Maximizing Patent Rules

We next examine optimal patent rules when the planner maximizes consumer surplus. In this case, the cost of innovation does not enter the PGA's objective function, so the PGA is only concerned about the duration of the race and the fraction of the benefit that consumers can retain. A reduction of the prize to the innovator increases consumer benefits, but slows down the arrival of the innovation. One way to relieve this tension is to use races to stimulate investment.

Figure 10 displays the optimal prize parameter γ^* and consumer surplus $W^C(\cdot)$ as a function of the cost ratio c for $\theta = 0$. Each line corresponds to a different D . The maximized consumer surplus is the upper envelope of the four lines in the figures.

Figure 10: Consumer surplus, $\theta = 0.0$



Several patterns are apparent in Figure 10. Consumer surplus decreases as the cost asymmetry rises. At small cost ratios the PGA can rely on the intense competition among the firms to ensure that the firms innovate quickly. Since the competition provides ample motivation for high investment levels, the PGA can set the prize-to-benefit ratio γ to be very low and the patent stage to $D = N = 5$. As c rises, the intensity of competition decreases since the inefficient firm reduces

investment. The PGA remedies this by increasing γ and by choosing a lower D . These changes spur both firms to work harder in the duopoly phase without creating too much risk that the inferior firm wins. In Figure 10, γ^* increases from 0.10 to 0.12 and D^* decreases from 5 to 2. As c increases further, even a short duopoly phase is not enough to motivate the firms. Since the PGA is reluctant to increase γ , the race becomes, for all practical purposes, just a monopoly innovation process by the more efficient firm. Thus the PGA is indifferent between setting D to any value between 1 to N .

Table 4 displays results for sensitivity analysis with respect to the parameters η , c and θ and confirms that the following result displayed in Figure 10 is robust to changes in θ and η : Consumer surplus is decreasing in c . Table 4 also reveals these additional results:

1. The optimal patent policy always involves a race, i.e. $D^* > 0$.
2. As in the social surplus case, γ^* is weakly decreasing in θ for a fixed c .
3. The optimal prize, γ^* , is always smaller than in the case of social surplus.
4. The optimal patent granting stage, D^* , is weakly decreasing in θ , regardless of c .

The intuition behind result (4), i.e. the pattern of the D^* values, is as follows. As the cost of investment for Firm 2 increases, its investment level declines; Firm 2 poses less of a competitive threat to Firm 1. In order to motivate both firms, the PGA lowers the optimal patent stage D^* , but this policy only partially motivates the firms to choose higher investment levels. When the cost of innovation is linear in investment effort and the cost ratio is sufficiently large, Firm 2 exits the race. Consequently, the probability of this firm advancing is zero, and the PGA is indifferent between all $D > 0$. When the cost function is strictly convex, Firm 2 never chooses a zero investment level since $C'(0) = 0$, and always has a positive chance of reaching $D^* = 1$ before Firm 1. As a result, D^* is always greater than 1. As in the social surplus case, this result highlights the only difference between strictly convex and linear cost.

5.3 Auctions for Allocating Patent Rights

Our previous results have shown that races arise endogenously in our environment. In this section, we enlarge the set of instruments available to the PGA by allowing it to choose an auction for innovation. The question is whether the PGA would prefer an auction to deliver the socially valuable product as opposed to a race and if yes, under what circumstances.

Suppose the PGA holds a second-price sealed-bid auction for the patent rights with a prize of $\Omega = \gamma B$. The PGA's payoff from an auction of patent rights before any investment or innovation

Table 4: Optimal Patent Policy for $B = 100, \beta = 0.996$

θ	c	$\eta = 1.5$			$\eta = 1.0$		
		D^*	γ^*	W^C	D^*	γ^*	W^C
0	1	5	0.10	80.7	4	0.10	81.0
	1.5	3	0.10	78.2	1	0.14	78.3
	2	3	0.12	76.4	*	0.12	77.9
	3	2	0.12	74.9	*	0.12	77.9
0.25	1	5	0.08	78.7	4	0.10	78.8
	1.5	3	0.10	76.0	*	0.10	75.3
	2	2	0.10	74.1	*	0.10	75.3
	3	2	0.10	72.3	*	0.10	75.3
0.4	1	5	0.08	77.7	4	0.10	77.4
	1.5	3	0.10	74.7	*	0.10	74.0
	2	2	0.10	72.8	*	0.10	74.0
	3	2	0.10	71.1	*	0.10	74.0
1	1	5	0.06	73.7	4	0.10	72.0
	1.5	3	0.08	70.6	*	0.10	68.9
	2	2	0.08	68.7	*	0.10	68.9
	3	2	0.08	66.3	*	0.10	68.9

* indicates that W^C is the same for all $D > 0$

takes place is the following:

$$(7) \quad P^{Auc}(\gamma; \alpha, C_i, C_j, \theta) = \max_{\gamma} \{W^{Auc}(i, 0) + V_j^M(0) + \alpha(V_i^M(0) - V_j^M(0))\},$$

where $W^{Auc}(i, 0)$ is the time $t = 0$ expected present discounted consumer surplus from the innovation when firm i is the innovator. It is formally defined as:

$$\begin{aligned} W^{Auc}(i, x_i) &= \beta \sum_{x'_i \geq x_i} p(x'_i | a_i^M(x_i), x_i) W^{Auc}(i, x'_i), \quad 0 \leq x_i < N, \\ W^{Auc}(i, N) &= (1 - \gamma)B - \theta\gamma B. \end{aligned}$$

The values $V_i^M(0)$ and $V_j^M(0)$ represent the private values of the patent to firm i and j respectively, and are equal to the monopolist inventors' values.

The parameter α is the Pareto weight placed on the firms, and is set to 0 when the PGA maximizes consumer surplus and to 1 in the case of social surplus maximization. In a second price auction where the winning firm pays the second bid, it is a well known result that the

dominant strategy for each firm is to bid exactly its expected net present value of the patent. In our environment, these private values are equal to the monopoly values for each firm. Under such a bidding strategy, the more efficient firm, Firm 1, would always win the patent. Therefore auctions can be very effective instruments for filtering out inefficient firms. However, when prizes are constrained, as in the case of reasonable θ values, auctions continue to be effective at filtering, but not so effective at spurring timely innovation. Races may dominate auctions in these cases.

Tables 5 and 6 report the optimal policy (including auctions as instruments) for a social surplus maximizing and consumer surplus maximizing PGA, respectively. Firms are assumed to be identical. The tables make it clear that given a choice between two mechanisms – auctions and patent races – the PGA would prefer auctions only when θ is low. When θ is of a more reasonable magnitude ($\theta \geq 0.25$ for social surplus and $\theta \geq 0.4$ for consumer surplus), however, auctions are dominated by races. The intuition behind this result is straightforward. In the auction, a monopolist innovator conducts all of the R&D and his rate of investment is below the socially efficient rate, especially when the prize is low due to higher θ . This leads to slower innovation and lowers the surplus from the auction. The patent race eases the trade off between the prize and the speed of innovation. The race itself provides incentives for higher investment and quicker innovation when the prize is not enough to motivate the firms. With a race of nontrivial duration, the PGA can achieve faster innovation than the auction can for the same prize. In these cases, the PGA prefers races.

Table 5: Social Surplus Maximizing Patent Policy for $B = 100$, $\beta = 0.996$, $c = 1$

θ	$\eta = 1.5$				$\eta = 1.0$			
	D^*	γ^*	P^{Auc}	policy	D^*	γ^*	P^{Auc}	policy
0.0		1.00	87.7	auction		1.00	86.3	auction
0.10		0.34	83.0	auction		0.26	82.4	auction
0.25	5	0.10	81.3	race	3	0.12	79.7	race
0.4	5	0.10	79.9	race	4	0.10	78.2	race
1.0	5	0.08	75.5	race	4	0.10	72.8	race

In sum, the results from the comparison of patent auctions vs. patent races once again underline the importance of considering deadweight losses generated in determining the best mechanism for innovation.

Table 6: Consumer Surplus Maximizing Patent Policy for $B = 100$, $\beta = 0.996$, $c = 1$

	$\eta = 1.5$				$\eta = 1.0$			
θ	D^*	γ^*	P^{Auc}	policy	D^*	γ^*	P^{Auc}	policy
0.0		1.00	87.7	auction		1.00	86.3	auction
0.10		0.34	83.0	auction		0.26	82.4	auction
0.25		0.22	79.4	auction		0.18	79.5	auction
0.4	5	0.08	77.7	race	4	0.10	77.4	race
1.0	5	0.06	73.7	race	4	0.10	72.0	race

6 Optimal Uniform Rules for Heterogeneous Innovations

In the previous sections, we assumed that the PGA knows the exact, ex-ante social value of the innovation, B . We also assumed that the PGA knows the cost technologies present, but can not assign them any particular firm. Therefore the PGA could tailor the patent rules to $c = \frac{c_2}{c_1}$ and B .¹⁴ In reality, patent laws are designed to apply across a broad range of industries and products, as well as a broad range of technologies and social benefits. To address this issue, we now consider the case where the PGA's information about the social value of the invention and the firms' technologies is restricted; it only knows the distributions of B and c , as well as their support, but not their exact values. Specifically, we assume that the PGA's beliefs about B are given by the probability density function $g(B)$, that its beliefs about c are represented by the density $f(c)$, and that $g(B)$ and $f(c)$ are independent. Given these beliefs and its social objective, the PGA maximizes the expected discounted social surplus, $\sum_{c,B} W^S(D, \gamma; \theta, B) f(c) g(B)$, or expected discounted consumer surplus $\sum_{c,B} W^C(D, \gamma; \theta, B) f(c) g(B)$.

In order to study this problem, we need to specify $f(c)$ and $g(B)$. We do not aim to execute a carefully calibrated exercise since the necessary data is not available. However, we do want to compute some "average" patent rules, illustrate the ease with which we can incorporate this into our analysis, and demonstrate the robustness of our previous results to this more general case. Therefore, we use the little data available to construct interesting examples. Pakes (1986) provides some documentation on the benefits of innovation for some European countries, and shows that their distribution is highly skewed: most innovations have very little or no social value and a few

¹⁴It may be possible to elicit information about a firm's costs. It may also be possible to hire firms to conduct R&D under the guidance of some central planner. However, that is not what a patent system does. Our analysis is a long way from being a fully specified mechanism design analysis; it represents instead the nature of feasible alternatives within a patent system. Our focus in this paper is on patent races, therefore we abstract from policies that would allow the PGA to conduct its own research and development by employing the firms in question.

innovations have very large values. However, these are ex-post realized values on innovations. Since the patent rules we analyze are chosen before any social value is realized, the relevant data for our model is the ex-ante distribution of values held by firms when they enter a patent race. Nevertheless, we assume that Pakes's empirical evidence represents an approximation for the distribution of ex-ante social values, so we use highly skewed distributions for B in our numerical results.

The first two rows of Table 7 display the supports for two distributions, B_1 and B_2 , of social values. The third row presents the probabilities for the possible values of B_i .

Table 7: Distribution of B

Support for B_1	10	32.5	57.5	85	120	160	210	277.5	380	600
Support for B_2	100	325	575	850	1200	1600	2100	2775	3800	6000
$\Pr(B)$	0.5	0.25	0.125	0.0625	0.0315	0.016	0.008	0.004	0.002	0.001

Because we have little data on innovation costs, we study optimal policies under four different cost distributions. As before, $c_1 = 1$. First, we examine two possible uniform distributions over a finite set of possible values denoted $Z = \{z_1, \dots, z_9\}$. The first, denoted U_1 assumes $z_i = 1 + (i-1).25$. The second, denoted U_2 assumes $z_i = i$. We look at both cases since they represent different degrees of heterogeneity in costs and different lack of information for the PGA. We do not want the results to strongly depend on the uniform specification. Therefore, we also consider two triangular distributions for c . More precisely, these distributions assume that the probability that $c = z_i$ is $(10 - i)/45$. We look at two possibilities for $Z = \{z_1, \dots, z_9\}$. The first, denoted T_1 , is $z_i = 1 + (i - 1).25$, and the second, denoted by T_2 , is $z_i = i$.

Table 8 reports our results for the social surplus maximizing policy when $\theta = 0$. The two rightmost columns represent the two possible beliefs about B , and the four bottom rows represent the four possible beliefs about c . As we move down the table, the mean and variance of the belief about c increases, and as we move right, the mean and variance of the belief about B increases. These results confirm the generality of our previous insights. When the PGA maximizes expected social surplus, and $\theta = 0$, the optimal policy has a bang-bang feature. The PGA chooses no race with full prize until sufficient variability in firms' costs is present. Then the optimal rule involves full length races with a smaller γ^* .

Table 9 reports the consumer surplus maximizing policy when $\theta = 0$. Consistent with our earlier results, the optimal policy always involves races and small γ^* . The results in Tables 8 and 9 present a few examples, but show that the results from the conditional analyses in Section 5 are robust to the more general case where the PGA must choose rules that apply over a wide variety of R&D processes.

Table 8: Social Surplus: Average Policies (D^*, γ^*) , $\theta = 0$

c	B_1	B_2
T_1	(0,1)	(0,1)
U_1	(0,1)	(0,1)
T_2	(5,0.38)	(0,1)
U_2	(5,0.40)	(5,0.16)

Table 9: Consumer Surplus: Average Policies (D^*, γ^*) , $\theta = 0$.

c	B_1	B_2
T_1	(3,0.18)	(5,0.02)
U_1	(3,0.20)	(5,0.04)
T_2	(3,0.20)	(3,0.06)
U_2	(2,0.20)	(3,0.06)

7 Extensions and Conclusions

Patent races are an integral part of the R&D process, but they do not represent the complete innovation process. A firm that has been granted a patent typically needs to incur additional costs and develop the product further before it can be produced and sold. We present an analysis of how the two parameters of the race – when the patent or exclusive contract is awarded and the winning prize – should be chosen in a simple multistage race.

We find that races of nontrivial duration are part of an optimal policy under most circumstances. In our setting, the patent race serves two purposes. First, it motivates the firms to invest and complete the innovation process quickly. When the prize causes inefficiencies, such as the monopoly grant implicit in a patent, using a race allows the planner to reduce the size of the prize and still give firms incentives to invest in innovation. Second, a race filters out inferior innovators since they cannot keep up with the more efficient ones.

The choice between short and long races depends on the social returns to innovation, the planner's objective (social vs. consumer surplus), and the inefficiency costs of compensating the patent winner. We show that in an environment with reasonably inefficient transfer mechanisms, longer races are preferred when firms are homogenous and shorter races are chosen otherwise. This result overturns the conventional wisdom that when firms are likely to compete fiercely, i.e., when they possess identical technologies in a simultaneous-move race, short races are preferable because they avoid excessive investment. Our analysis shows that this is true only when there is very little

constraint on the prizes a patent authority can give.

Our model allows us to understand the fundamental issues of developing a patent policy and identifying the complex trade-offs a patent authority faces. The environment we consider is a simple one, but our subsequent work indicates that the results are robust to many possible extensions. For example, one immediate extension is to consider races where firms can advance more than one stage at a time. We computed many such examples; they do not provide any substantial additional insights into the workings of the model. We have also studied cases where the technology of investment, i.e. the distribution F , depends on the stage of the innovation process. Again, no additional insights in terms of the trade-offs a patent authority faces were delivered by the modified technologies.

Another interesting extension is to allow firms to trade their technologies. It is straightforward to allow firms in our model to negotiate technology trades at each stage, similar to the trades examined in Green and Scotchmer (1995). In the context of our model, the technology leader may want to sell its technology to the laggard. We have studied this extension and found it to have no significant impact on the results for optimal patent policies.

Our results indicate that once a firm receives protection from competition, it reduces its investment level and slows the innovation process. The PGA varies the patent granting stage and the prize to induce firms to innovate quickly. In actual patent policy, there is a time limit on how long a product is protected under a patent. If firms develop the product too late, then they may not receive any (substantial) prize. This time limit could also serve both as a filtering device and an incentive for quick innovation, and therefore the planner may not rely on a race to differentiate between firms and spur investment. However, in all of the examples we computed, we chose parameters so that the time it takes for the firms to move from the patent-granting stage to the terminal innovation stage is short. Thus, the time limit of a patent would not significantly change any result.

It may be possible to devise other additional policy instruments that may remedy some of the inefficiencies that arise in the innovation race. One of the contributions of this paper is to identify the trade-offs the patent authority and firms face as the two fundamental features of patent policy – when a patent is granted and its associated prize – change, so that the choice of additional instruments is not made arbitrarily.

A Proofs

Proof of Proposition 1. We present the proof of this proposition for the case of strictly convex costs. The proof easily extends to the linear cost case, but it gets messy due to the possibility of corner

solutions. In the trivial case $\Omega = 0$ we have $V_i^M(x_i) = 0$ and $a^*(x_i) = 0$ for all $x_i \in \{0, 1, \dots, N\}$. Thus, we assume throughout the proof that $\Omega > 0$. The proof proceeds in four steps. First, we prove that there exists a solution to the Bellman equation. Second, we show that the value function is nondecreasing in the state. Third, we prove that there exists a unique optimal policy function. Finally, we show that the policy function is nondecreasing in the state.

Firm i 's monopoly problem is a dynamic programming problem with discounting that satisfies the standard assumptions for the existence of a solution, see Puterman (1994, Chapter 6) or Judd (1998, Chapter 12). The state space is finite. The discount factor satisfies $\beta < 1$. The cost function $C_i(\cdot)$ is continuous and thus bounded on the compact effort set A . The transition probability function $p(x_i'|\cdot, x_i)$ is also continuous on A for all $x_i \in \{0, 1, \dots, N\}$. Therefore, there exists a unique solution V_i^M to the Bellman equation and some optimal effort level $a^*(x_i)$ for each stage $x_i \in \{0, 1, \dots, N\}$.

Fix a state $x_i < N$ and an optimal effort level $a^*(x_i)$. The value $V_i^M(x_i)$ satisfies the equation

$$V_i^M(x_i) = \frac{-C_i(a^*(x_i)) + \beta p(x_i + 1|a^*(x_i), x_i)V_i^M(x_i + 1)}{1 - \beta p(x_i|a^*(x_i), x_i)}.$$

Since $C_i(\cdot)$ is nonnegative, $\beta < 1$, and $V_i^M(x_i + 1) \geq 0$ it follows that $V_i^M(x_i) \leq V_i^M(x_i + 1)$.

For the remainder of the proof we make use of the special form of the transition probability function p . Without loss of generality we assume that F is independent of the state x_i and write $F(x_i|x_i) = F < 1$. Under all our assumptions ($\Omega > 0$, $C(0) = 0$, $C'(0) = 0$, and $p(x_i|x_i, a_i) = F^{a_i}$) it holds that $V_i^M(x_i) > 0$ and $a^*(x_i) > 0$ for all $x_i \in \{0, 1, \dots, N\}$. Note that the optimal effort level is always in the interior of the set A . Given the value function V_i^M , a necessary (and sufficient) first-order condition for the optimal effort level is

$$F^a \beta \ln F (V_i^M(x_i) - V_i^M(x_i + 1)) - C_i'(a) = 0.$$

This equation must have a least one solution according to the first step of this proof. The second derivative of the function on the left-hand side equals $F^a \beta (\ln F)^2 (V_i^M(x_i) - V_i^M(x_i + 1)) - C_i''(a) < 0$. Hence, there is a unique optimal effort $a^*(x_i)$.

Given the value $V_i^M(x_i + 1)$, the optimal effort $a^*(x_i)$ and value $V_i^M(x_i)$ must be the (unique)

solution of the following system of two equations in the two variables a and V , respectively,

$$\begin{aligned} V(1 - \beta F^a) - \beta(1 - F^a)V_i^M(x_i + 1) + C(a) &= 0 \\ F^a \beta \ln F (V - V_i^M(x_i + 1)) - C'_i(a) &= 0 \end{aligned}$$

An application of the Implicit Function Theorem reveals that both variables in the solution are nondecreasing functions of the value $V_i^M(x_i + 1)$. The Jacobian of the function on the left-hand side at the solution equals

$$J = \begin{bmatrix} 1 - \beta F^a & 0 \\ F^a(\beta \ln F) & F^a \beta (\ln F)^2 (V - V_i^M(x_i + 1)) - C''(a) \end{bmatrix}.$$

The gradient of the function on the left-hand side with respect to the parameter $V_i^M(x_i + 1)$ equals

$$\begin{pmatrix} -\beta(1 - F^a) \\ -F^a \beta \ln F \end{pmatrix}.$$

The Implicit Function Theorem yields

$$\left(\frac{\frac{\partial V}{\partial V_i^M(x_i + 1)}}{\frac{\partial a}{\partial V_i^M(x_i + 1)}} \right) = -\frac{1}{D} \begin{bmatrix} F^a \beta (\ln F)^2 (V - V_i^M(x_i + 1)) - C''(a) & 0 \\ -F^a(\beta \ln F) & 1 - \beta F^a \end{bmatrix} \begin{pmatrix} -\beta(1 - F^a) \\ -F^a \beta \ln F \end{pmatrix} \geq 0,$$

where $D = (1 - \beta F^a)(F^a \beta (\ln F)^2 (V - V_i^M(x_i + 1)) - C''(a)) < 0$ is the determinant of the Jacobian. The value function V_i^M is nondecreasing in the state x_i and $a^*(x_i)$ is nondecreasing in the value $V_i^M(x_i + 1)$. Thus, the function a^* is nondecreasing in the state. \square

Proof of Theorem 1. We present again the proof for the case of strictly convex costs. For a given patent policy (D, γ) the strategy functions $\sigma_i^*, i = 1, 2$, constitute a Markov perfect equilibrium if they simultaneously solve equations (6). The proof is by backward induction. If $x_i = D$ for some i , then an optimal strategy pair $\sigma_i^*(x_i, x_{-i}), i = 1, 2$, and a pair of value functions $\mathbb{V}_i, i = 1, 2$, trivially exist. It is now sufficient to prove that for any state $(x_1, x_2) \in X$ with $x_i < D, i = 1, 2$, there exists a pure strategy Nash equilibrium (a_1^*, a_2^*) . To prove the existence of such an equilibrium we define a continuous function f on a convex and compact set such that any fixed point of this function is a pure strategy Nash equilibrium.

Given are a state $(x_1, x_2) \in X$ with $x_i < D, i = 1, 2$, and values $\mathbb{V}_i(x_i + 1, x_{-i}), \mathbb{V}_i(x_i, x_{-i} +$

1), $\mathbb{V}_i(x_i+1, x_{-i}+1)$ from the states that can be reached from (x_1, x_2) in one period. As in the proof of Proposition 1 we assume without loss of generality that the transition probability distribution is independent of the state and we write $F(x_i|x_i) = F$, $i = 1, 2$. We define a function f on a domain $S \equiv A \times [0, \gamma B] \times A \times [0, \gamma B]$. Choose an arbitrary element $(\hat{a}_i, V_i, \hat{a}_{-i}, V_{-i}) \in S$. Consider the equation

$$0 = -C'_i(a_i) \left(\frac{1}{F}\right)^{a_i} + \beta \ln F \cdot \left(F^{\hat{a}_{-i}}(V_i - \mathbb{V}_i(x_i+1, x_{-i})) + (1 - F^{\hat{a}_{-i}})(\mathbb{V}_i(x_i, x_{-i}+1) - \mathbb{V}_i(x_i+1, x_{-i}+1)) \right)$$

with the one unknown a_i . If $\delta \equiv F^{\hat{a}_{-i}}(V_i - \mathbb{V}_i(x_i+1, x_{-i})) + (1 - F^{\hat{a}_{-i}})(\mathbb{V}_i(x_i, x_{-i}+1) - \mathbb{V}_i(x_i+1, x_{-i}+1))$ is positive, then this equation has no solution. In this case we define $\acute{a}_i = 0$. If $\delta \leq 0$ then this equation has a unique solution $\acute{a}_i \geq 0$ (since $-C''_i(a_i) \left(\frac{1}{F}\right)^{a_i} + C'_i(a_i) \ln F \left(\frac{1}{F}\right)^{a_i} < 0$ for all $a_i \in A$). Note that $\acute{a}_i \in A$. We define $f_{i,1}(\hat{a}_i, V_i, \hat{a}_{-i}, V_{-i}) = \acute{a}_i$. Note that δ is continuous in V_i . An application of the Implicit Function Theorem shows that $f_{i,1}$ is continuous in V_i .

Next define \acute{V}_i by

$$\acute{V}_i = \frac{1}{1 - \beta F^{\acute{a}_i} F^{\hat{a}_{-i}}} \left(-C(\acute{a}_i) + \beta \left(F^{\acute{a}_i} (1 - F^{\hat{a}_{-i}}) \mathbb{V}_i(x_i, x_{-i}+1) + (1 - F^{\acute{a}_i}) F^{\hat{a}_{-i}} \mathbb{V}_i(x_i+1, x_{-i}) + (1 - F^{\acute{a}_i}) (1 - F^{\hat{a}_{-i}}) \mathbb{V}_i(x_i+1, x_{-i}+1) \right) \right).$$

Note that $\acute{V}_i \in [0, \gamma B]$ and define $f_{i,2}(\hat{a}_i, V_i, \hat{a}_{-i}, V_{-i}) = \acute{V}_i$. Clearly, the function $f_{i,2}$ is continuous.

In summary, we have defined a continuous function $f = (f_{1,1}, f_{1,2}, f_{2,1}, f_{2,2}) : S \rightarrow S$ mapping the convex and compact domain S into itself. Brouwer's fixed-point theorem implies that f has a fixed point $(a_1^*, V_1^*, a_2^*, V_2^*) \in S$. By construction of the function f this fixed point satisfies the equations (5) and (6). This completes the proof of the existence of a pure strategy Nash equilibrium in the state (x_1, x_2) . \square

B Computing Optimal Patent Policies

For any specific patent policy, (D, γ) , we need to compute the equilibrium of the race which involves solving two dynamic problems. First, we solve the dynamic optimization problem for each firm after it wins the patent. Second, we solve the patent race in the duopoly phase. We discuss the solution procedures for these two problems in detail.

B.1 Computing the Monopoly Phase

The monopoly phase begins after one of the firms reaches stage D , which can take any value between 0 and N . Therefore, we solve the monopoly problem for all $x_i \in \{0, 1, 2, \dots, N\}$, $i = 1, 2$. The successful firm's value function during the monopoly phase, V_i^M , solves the Bellman equation 2. We compute it by backward induction on states beginning at stage N and proceeding to the lower stages. At stage N , $V_i^M(N) = \Omega$ and $a_i^M(N) = 0$. Once we have computed $a_i^M(x')$ and $V_i^M(x')$ for $x' > x_i$, we can then compute the value functions $V_i^M(x_i)$ and policy functions $a_i^M(x_i)$ by using equations (2) and (3).

In addition to employing a standard value function iteration and implementing the Gauss-Seidel method for dynamic programming, (see p. 418 in Judd (1998)), we also occasionally use a second approach when the convergence criterion is very tight. This second approach solves a nonlinear system of first-order necessary and sufficient conditions. These conditions are necessary and sufficient given our assumption on the cost and Markov transition functions. The conditions are as follows:

$$(8) \quad V_i^M(x_i) = -C_i(a_i) + \beta \sum_{x'_i \geq x_i} p(x'_i | a_i, x_i) V_i^M(x'_i)$$

$$(9) \quad 0 = -C'_i(a_i) + \beta \sum_{x'_i \geq x_i} \frac{\partial}{\partial a_i} p(x'_i | a_i, x_i) V_i^M(x'_i) + \lambda_i$$

$$(10) \quad 0 = \lambda_i a_i$$

$$(11) \quad 0 \leq \lambda_i, a_i.$$

To find the solution to (8)–(11), we convert it into a nonlinear system of equations that guarantees a_i to be nonnegative. For this purpose we define

$$a_i = \max\{0, \alpha_i\}^\kappa \quad \text{and} \quad \lambda_i = \max\{0, -\alpha_i\}^\kappa$$

where $\kappa \geq 3$ is an integer and $\alpha_i \in \Re$. Note that, by definition, equation (10) and inequalities (11) are immediately satisfied. Thus, the unique solution to the nonlinear system of the two equations (8) and (9) with $a_i = \max\{\alpha_i, 0\}^\kappa$ in the two unknowns $V_i^M(x_i)$ and α_i yields the optimal policy and the corresponding value function of the monopolist.¹⁵

¹⁵The constraint on the effort level a can only be binding when the cost function C is linear. Nevertheless we use the constrained-optimization approach involving a Lagrange multiplier even when we use strictly convex cost functions. This approach is numerically much more stable than solving the first-order conditions of the unconstrained problem.

B.2 Solving the Duopoly Phase by an Upwind Procedure

The duopoly game has a finite set of states and could be solved using the techniques of Pakes and McGuire (1994). However, we have a special structure which allows for much faster computation. Since the game is over when one firm reaches D , the monopoly phase solution provides the value for each firm at all states (x_1, x_2) with $\max\{x_1, x_2\} = D$. The solution process for the remaining stages of the duopoly game utilizes a backward induction technique. For example, if we know the value at (D, D) , $(D - 1, D)$, and $(D, D - 1)$, then the game at $(D - 1, D - 1)$ reduces to a simple game where the only unknowns are the values and actions of each firm at $(D - 1, D - 1)$.

At each state (x_1, x_2) , we compute an equilibrium action pair $(\sigma_1(x_1, x_2), \sigma_2(x_1, x_2))$ and the corresponding values $(\mathbb{V}_1(x_1, x_2), \mathbb{V}_2(x_1, x_2))$ that satisfy conditions (5, 6). This computational task is surprisingly difficult; a Gauss-Seidel iterated best reply approach, a natural choice in such dynamic games that solves each firm's problem sequentially and updates their best responses to each other's actions, typically does not converge in our setting. Consequently we employ an alternative algorithm. We formulate the equilibrium problem in state (x_1, x_2) as a nonlinear system of equations. The following conditions are necessary and sufficient for optimality. For $i = 1, 2$,

$$(12) \quad 0 = -\mathbb{V}_i(x_i, x_{-i}) - C_i(a_i) + \beta \sum_{x'_i, x'_{-i}} p(x'_i | a_i, x_i) p(x'_{-i} | a_{-i}, x_{-i}) \mathbb{V}_i(x'_i, x'_{-i})$$

$$(13) \quad 0 = -\frac{\partial}{\partial a_i} C_i(a_i) + \beta \sum_{x'_i, x'_{-i}} \frac{\partial}{\partial a_i} p(x'_i | a_i, x_i) p(x'_{-i} | a_{-i}, x_{-i}) \mathbb{V}_i(x'_i, x'_{-i}) + \lambda_i$$

$$(14) \quad 0 = \lambda_i a_i$$

$$(15) \quad 0 \leq \lambda_i, a_i.$$

We transform this system of equations and inequalities into a nonlinear system of equations characterizing a Nash equilibrium at a state (x_1, x_2) with $x_i, x_{-i} < D$. We set $a_i = \max\{0, \alpha_i\}^\kappa$ and $\lambda_i = \max\{0, -\alpha_i\}^\kappa$ in equations (12) and (13) and omit the complementary slackness conditions (14) and the inequalities (15). The solutions to the resulting four nonlinear equations in the four unknowns $\mathbb{V}_i(x_i, x_{-i})$ and α_i for $i = 1, 2$, correspond to the Nash equilibrium of the stage game. Again we solve a constrained problem instead of an unconstrained problem since this choice results in a numerically much more stable procedure.

B.3 Optimal Patent Policy

The PGA maximizes its objective function W^S or W^C taking into consideration the effect of its policy (D, γ) on firms' investment. We parameterize the PGA's objective function in θ and B . Given the equilibrium strategies $\sigma_i(x)$ of the race and optimal policy function $a_i^M(x)$ during the monopoly phase, we can define the social surplus function W^S recursively as follows:

$$\begin{aligned}
W^{S,D}(x_1, x_2) &= -\sum_{i=1}^2 C_i(\sigma_i(x)) + \beta \sum_{x'_1, x'_2} p(x'_1 | \sigma_1(x), x_1) p(x'_2 | \sigma_2(x), x_2) W(x'_1, x'_2), \quad x_1, x_2 < D \\
W(x_1, x_2) &= \begin{cases} W^{S,D}(x_1, x_2), & x_1, x_2 < D \\ \frac{1}{2} (W^{S,M}(1, D) + W^{S,M}(2, D)), & x_1 = x_2 = D \\ W^{S,M}(i, x_i), & x_i = D \text{ and } x_{-i} < D, \quad i = 1, 2 \end{cases} \\
W^{S,M}(i, x_i) &= -C_i(a_i^M(x)) + \beta \sum_{x'_i \geq x_i} p(x'_i | a_i^M(x_i), x_i) W^{S,M}(i, x'_i), \quad x_i < N, \quad i = 1, 2 \\
W^{S,M}(N) &= B - \theta \gamma B.
\end{aligned}$$

The initial social surplus at $t = 0$ equals

$$W^S(D, \gamma; \theta, B) = W^{S,D}(0, 0).$$

The consumer surplus function W^C is similarly defined as

$$\begin{aligned}
W^{C,D}(x_1, x_2) &= \beta \sum_{x'_1, x'_2} p(x'_1 | \sigma_1(x), x_1) p(x'_2 | \sigma_2(x), x_2) W(x'_1, x'_2), \quad x_1, x_2 < D \\
W(x_1, x_2) &= \begin{cases} W^{C,D}(x_1, x_2), & x_1, x_2 < D \\ \frac{1}{2} (W^{C,M}(1, D) + W^{C,M}(2, D)), & x_1 = x_2 = D \\ W^{C,M}(i, x_i), & x_i = D \text{ and } x_{-i} < D, \quad i = 1, 2 \end{cases} \\
W^{C,M}(i, x_i) &= \beta \sum_{x'_i \geq x_i} p(x'_i | a_i^M(x_i), x_i) W^{C,M}(i, x'_i), \quad x_i < N, \quad i = 1, 2 \\
W^{C,M}(N) &= (1 - \gamma)B - \theta \gamma B.
\end{aligned}$$

Initial consumer surplus at $t = 0$ equals

$$W^C(D, \gamma; \theta, B) = W^{C,D}(0, 0).$$

Definition 2. *The social surplus maximizing patent policy is a pair (D^*, γ^*) that maximizes*

$W^S(D, \gamma; \theta, B)$ given (θ, B) . The consumer surplus maximizing patent policy is a pair (D^*, γ^*) that maximizes $W^C(D, \gamma; \theta, B)$ given (θ, B) .

We solve the dynamic equilibrium of the patent race for a large discrete set of (D, γ) pairs to find the optimal PGA policy (D^*, γ^*) . The ratio γ takes values from a discrete set $\Gamma \subset [0, \bar{\gamma}]$. We summarize all computational steps in the following algorithm.

Algorithm 1 (Computation of welfare-maximizing policy).

1. Select an objective function $W \in \{W^S, W^C\}$. Fix the parameters θ and B . Choose a grid $\Gamma \subset [0, 1]$.
2. For each $\gamma \in \Gamma$
 - (a) Set $\Omega = \gamma B$.
 - (b) Solve the monopoly problem given Ω .
 - (c) For $D = 0$, compute the expected planner surplus, $W(0, \gamma; \theta, B)$, of giving the patent monopoly to a firm chosen randomly with equal probabilities.
 - (d) For each $D \in \{1, 2, \dots, N\}$
 - i. Solve the duopoly game for $x_1, x_2 < D$.
 - ii. Compute the expected planner surplus, $W(D, \gamma; \theta, B)$
3. Find the optimal (D^*, γ^*) which maximizes $W(D, \gamma; \theta, B)$.

References

- [1] Dasgupta, P. and Stiglitz, J. , “Industrial Structure and the Nature of Innovative Activity,” Economic Journal 90 (1980a), 266–293.
- [2] Dasgupta, P. and Stiglitz, J., “Uncertainty, Market Structure and the Speed of Research,” Bell Journal of Economics 11 (1980b), 1–28.
- [3] Denicolo, V., “The Optimal Life of a Patent when Timing of Innovation is Stochastic,” International Journal of Industrial Organization 17 (1999), 827–846.
- [4] Denicolo, V., “Two-stage patent races and patent policy,” RAND Journal of Economics, 31 (2000), 488–501.

- [5] Doraszelski, U., “An R&D Race with Knowledge Accumulation,” Rand Journal of Economics, 34 (2003), 19–41.
- [6] Fershtman, C. and Markovich, S., “Patents, Imitation, and Licensing in an Asymmetric Dynamic R&D Race,” International Journal of Industrial Organization 28 (2010), 113–126.
- [7] Fudenberg, D., Gilbert, R., Stiglitz, J. and Tirole, J., “Preemption, Leapfrogging, and Competition in Patent Races,” European Economic Review 22 (1983), 3–31.
- [8] Fullerton, R. and McAfee, P., “Auctioning Entry into Tournaments,” Journal of Political Economy 107 (1999), 573–605.
- [9] Gilbert, R. and Shapiro, C., “Optimal Patent Length and Breadth,” RAND Journal of Economics, 21 (1990), 106–112.
- [10] Green, J. and Scotchmer, S., “Novelty and Disclosure in Patent Law,” RAND Journal of Economics 21 (1990), 131–146.
- [11] Green, J. and Scotchmer, S., “On the Division of Profit in Sequential Innovation,” RAND Journal of Economics 26 (1995), 20–33.
- [12] Harris, C. and Vickers, J., “Perfect Equilibrium in a Model of a Race,” Review of Economic Studies 52 (1985a), 193–209.
- [13] Harris, C. and Vickers, J., “Patent Races and the Persistence of Monopoly,” Journal of Industrial Economics 33 (1985b), 461–481.
- [14] Harris, C. and Vickers, J., “Racing with Uncertainty,” Review of Economic Studies, 54 (1987), 1–21.
- [15] Hopenhayn H., Llobet G. and Mitchell, M. “Rewarding Sequential Innovators: Patents, Prizes, and Buyouts,” Journal of Political Economy 114 (2006), 1041 – 1068.
- [16] Hopenhayn, H. and Mitchell, M. “Innovation Variety and Patent Breadth,” The RAND Journal of Economics 32 (2001), 152–166.
- [17] Jewkes, J., Sawers, D., and Stillerman, R. The Sources of Invention, (New York, N.Y: Norton, 1969).
- [18] Judd, K. L., “Closed-loop Equilibrium in an Innovation Race with Experience,” mimeo, Northwestern University, 1985a.

- [19] Judd, K. L., “Closed-Loop Equilibrium in a Multistage Innovation Race,” mimeo, The Center for Mathematical Studies in Economics, Northwestern University, 1985b.
- [20] Judd, K. L. “The Welfare Cost of Factor Taxation in a Perfect Foresight Model,” *Journal of Political Economy*, 95, 675–709.
- [21] Judd, K. L., Numerical Methods in Economics. (Cambridge, MA: MIT Press, 1998)
- [22] Kamien, M. and Schwartz, N., Market Structure and Innovation, (New York, NY: Cambridge University Press, 1982)
- [23] Klemperer, P., “How Broad should the Scope of a Patent Protection Be?” RAND Journal of Economics 21 (1990), 113–130.
- [24] Lee, T. and Wilde, L., “Market Structure and Innovation: A Reformulation,” *Quarterly Journal of Economics*, 94 (1980), 429–436.
- [25] Loury, G., “Market Structure and Innovation,” The Quarterly Journal of Economics 93 (1979), 395–410.
- [26] Nordhaus, W. D., Invention, Growth and Welfare (Cambridge: M.I.T. Press, 1969)
- [27] O’Donoghue, T., Scotchmer, S. and Thisse, J-F. “Patent Breadth, Patent Life, and the Pace of Technological Progress,” Journal of Economics, Management and Strategy 7 (1998), 1–32.
- [28] Pakes, A., “Patents as Options: Some Estimates of the Value of Holding European Patent Stocks.” Econometrica, 54 (1986), 755–784.
- [29] Pakes, A. and McGuire, P., “Computing Markov-perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model,” RAND Journal of Economics 25 (1994), 555–589.
- [30] Perry, M. and Vincent, D., “The Optimal Timing of Procurement Decisions and Patent Allocations,” International Economic Review, 43 (2002), 1035–1052.
- [31] Puterman, M. L., Markov Decision Processes: Discrete Stochastic Dynamic Programming, (New York, NY: John Wiley & Sons, Inc., 1994)
- [32] Reinganum, J., “Dynamic Games of Innovation,” Journal of Economic Theory 25 (1981), 21–41.

- [33] Reinganum, J., “A Dynamic Game of R&D: Patent Protection and Competitive Behaviour,” Econometrica, 50 (1982), 671–688.
- [34] Reinganum, J., “The Timing of Innovation: Research, Development, and Diffusion,” in R. Schmalensee and R. D. Willig, eds. Handbook of Industrial Organization 1 (Amsterdam: The Netherlands: Elsevier, 1989)
- [35] Scotchmer, S., “Protecting Early Innovators: Should Second Generation Products be Patentable?” RAND Journal of Economics 27 (1996) , 117–126.
- [36] Scotchmer, S., “On the Optimality of the Patent Renewal System,” RAND Journal of Economics 30 (1999), 181–196.
- [37] Taylor, C., ”Digging for Golden Carrots: An Analysis of Research Tournaments,” American Economic Review 85 (1995), 872–889.
- [38] Wright, B. D., “The Economics of Invention Incentives: Patents, Prizes and Research Contracts,” American Economic Review 73 (1983), 691–707.
- [39] Wright, D. J., “Optimal Patent Breadth and Length with Costly Imitation,” International Journal of Industrial Organization 17 (1999), 419–436.